

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:47 ; Search time 91.75 Seconds

(without alignments)
15.135 Million cell updates/sec

Title: US-09-331-631a-5_COPY_33_75

Perfect score: 248
Sequence: 1 NOEDPQTECQOCQRRCKROOE.....RQOYCORCKEICEEEREY 43

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 105 | 42.3 | 588 | 1 | VCIA_GOSHI |
| 2 | 94 | 37.9 | 605 | 1 | VCIA_GOSHI |
| 3 | 71 | 28.6 | 47 | 1 | AGRP_LOFCY |
| 4 | 68 | 27.4 | 154 | 1 | YS1L_CAEEL |
| 5 | 65.5 | 26.4 | 33 | 1 | MBPI_MAIZE |
| 6 | 65 | 26.2 | 919 | 1 | ANDR_HUMAN |
| 7 | 64 | 25.8 | 648 | 1 | KAPC_DICDI |
| 8 | 64 | 25.8 | 911 | 1 | ANDR_PANTR |
| 9 | 63.5 | 25.6 | 50 | 1 | HSP1_MOUSE |
| 10 | 63.5 | 25.6 | 50 | 1 | HSP1_MOUSE |
| 11 | 63 | 25.4 | 626 | 1 | AH12_ARAHY |
| 12 | 61 | 24.6 | 223 | 1 | CAS2_SHEEP |
| 13 | 61 | 24.6 | 429 | 1 | AP4A_MACFA |
| 14 | 61 | 24.6 | 907 | 1 | ANDR_CANFA |
| 15 | 60 | 24.0 | 285 | 1 | INVO_CANFA |
| 16 | 59.5 | 24.0 | 614 | 1 | AH11_ARAHY |
| 17 | 59 | 23.8 | 758 | 1 | YM38_YEAST |
| 18 | 59 | 23.8 | 2124 | 1 | Y192_HUMAN |
| 19 | 58.5 | 23.6 | 639 | 1 | GLCX_SOYBN |
| 20 | 58.5 | 23.6 | 1407 | 1 | TRHX_RABIT |
| 21 | 58 | 23.4 | 223 | 1 | CAS2_CAPHI |
| 22 | 58 | 23.4 | 644 | 1 | BTJL_DROME |
| 23 | 58 | 23.4 | 2318 | 1 | NTCA_MOUSE |
| 24 | 57.5 | 23.2 | 338 | 1 | FSH_HUMAN |
| 25 | 57.5 | 23.2 | 605 | 1 | GLCA_SOYBN |
| 26 | 57.5 | 23.2 | 991 | 1 | DHPI_SCHPO |
| 27 | 57 | 23.0 | 210 | 1 | YOEK_ECOLI |
| 28 | 57 | 23.0 | 1898 | 1 | TRHX_HUMAN |
| 29 | 57 | 23.0 | 1905 | 1 | TAGB_DICDI |
| 30 | 56.5 | 22.8 | 284 | 1 | TPM1_RAT |
| 31 | 56.5 | 22.8 | 284 | 1 | TPM2_HUMAN |
| 32 | 56.5 | 22.8 | 284 | 1 | TPMB_HUMAN |
| 33 | 56.5 | 22.8 | 284 | 1 | TPMB_RABIT |

| | | | | | |
|----|------|------|-----|---|------------|
| 34 | 56 | 22.6 | 239 | 1 | CAUD_MELGA |
| 35 | 56 | 22.6 | 524 | 1 | SBP_SOYBN |
| 36 | 56 | 22.6 | 771 | 1 | CAUD_CHICK |
| 37 | 55.5 | 22.4 | 49 | 1 | HSP1_PIG |
| 38 | 55.5 | 22.4 | 174 | 1 | GRE_METTI |
| 39 | 55.5 | 22.4 | 284 | 1 | TPM1_CHICK |
| 40 | 55.5 | 22.4 | 284 | 1 | TPM1_MOUSE |
| 41 | 55.5 | 22.4 | 284 | 1 | TPMB_CHICK |
| 42 | 55.5 | 22.4 | 303 | 1 | BNK_DROME |
| 43 | 55.5 | 22.4 | 441 | 1 | PPAS_HUMAN |
| 44 | 55 | 22.2 | 339 | 1 | TF2D_HUMAN |
| 45 | 55 | 22.2 | 450 | 1 | SFR_DROME |

ALIGNMENTS

| RESULT | ID | VCIA_GOSHI | STANDARD | PRT | 588 AA |
|--------|--|------------|----------|-----|--------|
| AC | P09801 | VCIA_GOSHI | STANDARD | PRT | 588 AA |
| DT | 01-MAR-1989 (Rel. 10, Created) | | | | |
| DT | 01-MAR-1989 (Rel. 10, Last sequence update) | | | | |
| DT | 15-JUL-1999 (Rel. 38, Last annotation update) | | | | |
| DE | VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B). | | | | |
| OS | Gossypium hirsutum (Upland cotton). | | | | |
| OC | Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium. | | | | |
| RN | [1] | | | | |
| RA | SEQUENCE FROM N.A. | | | | |
| RP | Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III; | | | | |
| RT | "Developmental biochemistry of cottonseed embryogenesis and germination. XVII. cDNA and amino acid sequences of the members of the storage protein families." | | | | |
| RL | Plant Mol. Biol. 7:475-489(1986). | | | | |
| CC | -1- FUNCTION: SEED STORAGE PROTEIN. | | | | |
| CC | -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN BODIES. | | | | |
| CC | -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICITIN, ETC.). | | | | |
| CC | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch). | | | | |
| DR | EMBL; M16891; AAA33071.1; - | | | | |
| DR | PIR; A30838; FMCNAB. | | | | |
| DR | HSSP; P50477; ICAX. | | | | |
| DR | INTERPRO; IPR001113; - | | | | |
| DR | PFAM; PF00546; Seedstore_7s; 1. | | | | |
| KW | Seed storage protein; Signal. | | | | |
| FT | SIGNAL | | | | |
| FT | CHAIN | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |
| FT | 588 AA; 69729 MW; 63699929AB8ADEB CRC64; | | | | |
| FT | VICILIN C72. | | | | |
| FT | SEQUENCE | | | | |

```

AC P09799;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borrieto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RC Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s. 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 37.9%; Score 94; DB 1; Length 605;
Best Local Similarity 35.9%; Pred. No. 0.0074;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

OY 2 QEDPTECCQCORRQOESDPROOQYCORCKEICEE 40
DB 115 QQQPKQFKKCCQRCQWQEQREPRKQCYKCEKREQYQED 153

RESULT 3
ACRP_LUPCY STANDARD; PRT; 47 AA.
ID AGRP_LUPCY
AC P56568;
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE 6.5 KDA ARGININE/GLUTAMATE-RICH POLYPEPTIDE (6.5K-AGRP).
OS Luffa cylindrica (Smooth loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
RP SEQUENCE.
RC TISSUE=SEED;
RX MEDLINE=97357433; PubMed=9214759;
RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;
RT "Primary structure of 6.5k-arginine/glutamate-rich polypeptide from
RT the seeds of sponge gourd (Luffa cylindrica).";
RL Biosci. Biotechnol. Biochem. 61:984-988(1997).
CC -1- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
CC RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC -1- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.

```

```

CC -1- SIMILARITY: SOME TO 7S SEED STORAGE PROTEINS.
KW Seed storage protein.
FT DISULFID 12 33
FT DISULFID 16 29
SQ SEQUENCE 47 AA; 5698 MW; 588B0EC82273AC05 CRC64;

Query Match 28.6%; Score 71; DB 1; Length 47;
Best Local Similarity 37.1%; Pred. No. 0.18;
Matches 13; Conservative 9; Mismatches 9; Indels 4; Gaps 1;

OY 5 PTECCQCCORRQOESDPROOQYCORCKEICEE 39
DB 5 PRTEYACRVRQVAHGVHVER---QRCQOVCCK 35

RESULT 4
Y5S1_CAEEL STANDARD; PRT; 154 AA.
ID Y5S1_CAEEL
AC Q09662;
DT 01-NOV-1995 (rel. 32, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE HYPOTHETICAL 15.6 KDA PROTEIN ZK673.1 IN CHROMOSOME II PRECURSOR.
GN ZK673.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RC McKurray A.;
RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP REVISIONS.
RC STRAIN=BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO C.ELEGANS E04D5.4 AND SOME, TO C.ELEGANS
CC C03G6.13.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z48585; CAA84846.1; -.
DR HSSP: P01551; IACX.
DR WORMPEP: ZK673.1; CE16742.
KW Hypothetical protein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 154 POTENTIAL.
SQ SEQUENCE 154 AA; 15637 MW; 323A02C3A2E782AB CRC64;

Query Match 27.4%; Score 68; DB 1; Length 154;
Best Local Similarity 35.5%; Pred. No. 0.96;
Matches 11; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

OY 3 EDPTECCQCORRQOESDPROOQYCORCK 33
DB 69 DDPNTDCITQYTLGSAKTYPILOQFCPRTC 99

RESULT 5
MBP1_MAIZE STANDARD; PRT; 33 AA.
ID MBP1_MAIZE
AC P28734;
DT 01-DEC-1992 (rel. 24, Created)
DT 01-DEC-1992 (rel. 24, Last sequence update)

```

DT 01-OCT-1996 (Rel. 34, last annotation update)
DE ANTIMICROBIAL PEPTIDE MBP-1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE.
RC STRAIN=CV. B73; TISSUE=SEED;
RX MEDLINE=92406801; Pubmed=1527010;
RA DUVICK J. P., Rood T., Rao A. G., Marshak D. R.;
RT "Purification and characterization of a novel antimicrobial peptide
from maize (Zea mays L.) kernels.";
RL J. Biol. Chem. 267:18814-18820(1992).
CC -1- FUNCTION: INHIBITOR OF BOTH BACTERIAL AND FUNGAL GROWTH IN VITRO
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE EMBRYO PORTION OF THE
CC KERNEL.
DR PIR: A41822; A41822.
DR MATZEDB: 69182; -.
KW Fungicide; Antidiabetic.
SQ SEQUENCE 33 AA: 4131 MW: B148F19B0E823599 CRC64;

| | | | | |
|-----------------------|-----------------|-----------------|-----------|------------|
| Query Match | 26.4%; | Score 65.5; | DB 1; | length 33; |
| Best Local Similarity | 44.0%; | Pred. No. 0.49; | | |
| Matches 11; | Conservative 6; | Mismatches 7; | Indels 1; | Gaps 1 |

```
Oy      11 QcQRC-RQESDPRQQYcQRCK 34  
        :|:|:| | | | | |  
Db      6 ECRQCRLRRHGGQPWETQECMRRCR 30
```

| RESULT | 6 | | | | |
|------------|---|-----------|------|-----|-----|
| ANDR_HUMAN | ANDR_HUMAN | STANDARD; | PRT; | 919 | AA. |
| AC | P10275; | | | | |
| DT | 01-MAR-1989 (Rel. 10, Created) | | | | |
| DT | 01-APR-1990 (Rel. 14, Last sequence update) | | | | |
| DT | 01-OCT-2000 (Rel. 40, Last annotation update) | | | | |
| DE | ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR). | | | | |
| GN | AR OR NR3CA OR DHTR. | | | | |
| OS | Homo sapiens (Human). | | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=89112208; PubMed=3216866; | | | | |
| RA | Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E., | | | | |
| RA | French F.S., Wilson E.M.; | | | | |
| RT | "The human androgen receptor: complementary deoxyribonucleic acid | | | | |
| RT | cloning, sequence analysis and gene expression in prostate."; | | | | |
| RL | Mol. Endocrinol. 2:1265-1275(1988). | | | | |
| RN | [2] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=90083302; PubMed=2594783; | | | | |
| RA | Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J., | | | | |
| RA | Wilson E.M., French F.S.; | | | | |
| RT | "Sequence of the Intron/exon junctions of the coding region of the | | | | |
| RT | human androgen receptor gene and identification of a point mutation | | | | |
| RT | in a family with complete androgen insensitivity."; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989). | | | | |
| RN | [3] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RX | MEDLINE=90258935; PubMed=2342476; | | | | |
| RA | Govindan M.V.; | | | | |
| RT | "Specific region in hormone binding domain is essential for hormone | | | | |
| RT | binding and trans-activation by human androgen receptor."; | | | | |
| RL | Mol. Endocrinol. 4:417-427(1990). | | | | |
| RN | [4] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RP | TISSUE=PROSTATE; | | | | |
| RX | MEDLINE=89017168; PubMed=3174628; | | | | |
| RA | Chang C., Kokontis J., Lao S.; | | | | |
| RT | "Structural analysis of complementary DNA and amino acid sequences of | | | | |

human and rat androgen receptors.";
Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
[5]
SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcello M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen
RL receptor.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
[6]
SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=9115943; PubMed=2293020;
RA Marcello M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RT McPhaul M.J.;
RN "Definition of the human androgen receptor gene structure permits the
RT identification of mutations that cause androgen resistance: premature
RT termination of the receptor protein at amino acid residue 588 causes
RL complete androgen resistance.";
RN Mol. Endocrinol. 4:1105-1116(1990).
[7]
SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=8817811; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding
RN androgen receptors.";
RL Science 240:324-326(1988).
[8]
SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=8824040; PubMed=337788;
RA Trippan J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RT Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RN Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human
RN androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
[9]
POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledgens H.F., Oostera B.A., Brinkmann A.O., Trippan J.;
RT "Trinucleotide repeat polymorphism in the androgen receptor gene
RN (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
[10]
POLYMORPHISM OF POLY-GLY REGION.
RC TISSUE=BL00D;
RX Lu J., Danielson M.;
RN Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
[11]
VARIANTS SBMA IN POLY-GLN REGION.
RX MEDLINE=91287825; PubMed=2052380;
RA la Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischback K.H.;
RT "Androgen receptor gene mutations in X-linked spinal and bulbar
RN muscular atrophy.";
RL Nature 352:77-79(1991).
[12]
REVIEW ON VARIANTS.
RX MEDLINE=95023089; PubMed=7937057;
RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RN Nucleic Acids Res. 22:3560-3562(1994).
[13]
REVIEW ON VARIANTS.
RX MEDLINE=97109385; PubMed=9016528;
RA Gottlieb B., Trifiro M., Lumbroso R., Vasilou D.M., Pinsky L.;
RT "The androgen receptor gene mutations database.";
RN Nucleic Acids Res. 25:158-162(1997).
[14]
VARIANT LNCAP ALA-877.
RX MEDLINE=91083633; PubMed=2260966;
RA Veldscholte J., Ris-Stalpers C., Kuiper G.G., Jenster G.,
RN Bervoets C., Klaassen B., van Rooij H.C.J., Trippan J.

| | |
|----|--|
| CC | entities requires a license agreement (see http://www.lsb-sib.ch/announce/ |
| CC | or send an email to license@lsb-sib.ch . |
| CC | ----- |
| DR | EMBL; M38703; -, NOT_ANNOTATED_CDS. |
| DR | PIR; J01150; J01150. |
| DR | HSSP; P05132; ZCPK. |
| DR | DICTYDB; DD02030; PKAC. |
| DR | INTERPRO; IPR000719; -. |
| DR | INTERPRO; IPR000961; -. |
| DR | INTERPRO; IPR002290; -. |
| DR | PRAM; PR00069; kinase; 1. |
| DR | PRAM; PF00433; kinase; 1. |
| DR | PROSITE; PS00107; PROTEIN_KINASE_ATP; 1. |
| DR | PROSITE; PS00108; PROTEIN_KINASE_ST; 1. |
| DR | PROSITE; PS50011; PROTEIN_KINASE_DOM; 1. |
| KW | Transferase; Serine/threonine-protein kinase; ATP-binding; CAMP; |
| KW | Phosphorylation. |
| FT | DOMAIN 58 64 ASN-RICH. |
| FT | DOMAIN 136 223 GLN-RICH. |
| FT | DOMAIN 233 250 THR-RICH. |
| FT | DOMAIN 336 590 PROTEIN KINASE. |
| FT | NP_BIND 342 350 ATP (BY SIMILARITY). |
| FT | BLINDING 365 365 ATP (BY SIMILARITY). |
| FT | ACT_SITE 459 459 BY SIMILARITY. |
| FT | MOD_RES 490 490 PHOSPHORYLATION (BY SIMILARITY). |
| SQ | SEQUENCE 648 AA: 74458 MW: 165934 Da GC64; |

| | | | | | |
|----|---|--------|----------------|-----------|-------------|
| | Query Match | 25.8% | Score 64; | DB 1; | Length 648; |
| | -Best Local Similarity | 29.3%; | Pred. No. 7.7; | | |
| | Matches 12; Conservative | 15; | Mismatches 14; | Indels 0; | Gaps 0; |
| QY | 2 QEDPQTECQCQCRCRCRQESDPQQQVYQRCKETCEELEE 42 | | | | |
| | : : : : : : : : : : : : : : | | | | |
| bB | 145 QQQPQQQQPPQQQQPPQQQQQQPPQQQQQLQQNNNNQQQQ 185 | | | | |

[illegible]

| | |
|----|---|
| DR | EMBL: U94177; AAC73048.1; -. |
| DR | HSSP: P06536; IRGD |
| DR | INTERPRO: IPR000536; -. |
| DR | INTERPRO: IPR001103; -. |
| DR | INTERPRO: IPR001628; -. |
| DR | PFAM: PF00104; hormone_rec; 1. |
| DR | PFAM: PF00105; zf-C4; 1. |
| DR | PRINTS: PR00047; STROIDFINGER. |
| DR | PRINTS: PR00521; ANDROGENR. |
| DR | PROSITE: PS00031; NUCLEAR_RECEPTOR; 1. |
| KW | Receptor; Transcription regulation; DNA-binding; Nuclear protein; |
| RK | Zinc-finger; Steroid-binding. |
| FT | DOMAIN 1 549 MODULATING (BY SIMILARITY). |
| FT | DNA_BIND 551 616 C4-TYPE ZINC FINGERS (TWO). |
| FT | ZN_FING 551 571 C4-TYPE. |
| FT | ZN_FING 587 611 C4-TYPE. |
| FT | DOMAIN 682 911 LIGAND-BINDING. |
| FT | DOMAIN 57 78 POLY-GLN. |
| FT | DOMAIN 84 88 POLY-GLN. |
| FT | DOMAIN 192 196 POLY-GLN. |
| FT | DOMAIN 371 380 POLY-PRO. |
| FT | DOMAIN 395 401 POLY-ALA. |
| FT | POLY-GLY 448 464 |
| SQ | SEQUENCE 911 AA; 98402 MW; 601B9BD4E697DAA4 CRC64; |

```
Query Match          25.8%; Score 64; DB 1; Length 911;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 13; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
```

| RESULT 9 | | | |
|----------|--|-----------|-------------|
| ID | HSPL_MOUSE | STANDARD: | PRT: 50 AA. |
| AC | P02319; | | |
| DT | 21-JUL-1986 (Rel. 01, Created) | | |
| DT | 21-JUL-1986 (Rel. 01, Last sequence update) | | |
| DT | 15-DEC-1988 (Rel. 37, Last annotation update) | | |
| DE | SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE). | | |
| GN | PRM1 OR PRM-1. | | |
| OS | Mus musculus (Mouse). | | |
| OC | Eukaryota; Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=85199803; Pubmed=2986684; | | |
| RA | Kleene K.C., Distel R.J., Hecht N.B.; | | |
| RT | "Nucleotide sequence of a cDNA clone encoding mouse protamine 1."; | | |
| RL | Biochemistry 24:719-722(1985). | | |
| RN | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=88193085; Pubmed=3358932; | | |
| RA | Johnson P.A., Paschon J.J., Yelick P.C., Palminter R.D., Hecht N.B.; | | |
| RT | "Sequence homologies in the mouse protamine 1 and 2 genes."; | | |
| RL | Biochem. Biophys. Acta 950:45-53(1988). | | |
| RN | [3] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=87260978; Pubmed=3037541; | | |
| RA | Paschon J.J., Behringer R.R., Brinster R.L., Palminter R.D.; | | |
| RT | "Spermatid-specific expression of protamine 1 in transgenic mice."; | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 84:5316-5319(1987). | | |
| RN | [4] | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=88181903; Pubmed=3445973; | | |
| RA | Hecht N.B.; | | |
| RT | "Gene expression during spermatogenesis."; | | |
| RL | Ann. N.Y. Acad. Sci. 513:90-101(1987). | | |
| RN | [5] | | |
| RP | SEQUENCE FROM N.A. | | |

```

RC STRAIN=C129;
RA Schlueter G., Engel W.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
CC -!- FUNCTION: PROTAMINE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC Sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
CC DNA-HELIX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02926; AAA39980.1; -
DR EMBL: X07625; CAA30472.1; -
DR EMBL: X14003; CAA32169.1; -
DR EMBL: M27500; AAA39985.1; -
DR EMBL: Z47352; CAA87410.1; -
DR PIR: A02660; HSMMS1.
DR PIR: S03820; S03820.
DR PIR: A28331; A28331.
DR MGD: MGI:97765; PRL1.
DR INTERPRO: IPR000221; -
DR PFAM: PF00260; protamine.P1.1.
DR PROSITE: PS00048; protamine.P1.1.
DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MER 0 0
FT DISULFID 5 5 INTERCHAIN (WITH CYS-21) (BY SIMILARITY).
FT DISULFID 6 14 BY SIMILARITY.
FT DISULFID 21 21 INTERCHAIN (WITH CYS-5) (BY SIMILARITY).
FT DISULFID 36 36 INTERCHAIN (WITH CYS-36) (BY SIMILARITY).
FT DISULFID 37 47 BY SIMILARITY.
SQ SEQUENCE 50 AA: 6827 MW: B61CAF3D1BBFA978 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 50;
Best Local Similarity 28.9%; Pred. No. 1.1;
Matches 11; Conservative 11; Mismatches 11; Indels 5; Gaps 1;

OY 6 QTECCQOCRRHQESDPPOQCYCQRKEICEEEY 43
DB 11 RSRRCRRRRRCRR-----RRRCRRRRRRRCRRRSY 43

RESULT 10
HSP1_RAT STANDARD: PRT; 50 AA.
ID HSP1_RAT
AC P10118:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPERM PROTAMINE P1 (CYSTEINE-RICH PROTAMINE).
GN PRM1 OR PRM-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89335257; PubMed=2757789;
RA Klemm U., Lee C.H., Butteind P., Hake S., Engel W.;
RT "Nucleotide sequence of a cDNA encoding rat protamine and the haploid
RT expression of the gene during rat spermatogenesis.";
RN Biol. Chem. Hoppe-Seyler 370:293-301(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96341725; PubMed=8720108;

```

```

RA Schlueter G., Celik A., Obata R., Schlicker M., Hofferbert S.,
RA Schluger A., Adham I.M., Engel W.;
RT "Sequence analysis of the conserved protamine gene cluster shows that
RT it contains a fourth expressed gene.";
RN Mol. Reprod. Dev. 43:116(1996).
RN [3]
RP SEQUENCE.
RX MEDLINE=89207111; PubMed=3072011;
RA Ammer H., Henschen A.;
RT "Rat sperm protamine. Isolation and sequence analysis.";
RN Biol. Chem. Hoppe-Seyler 369:1301-1306(1988).
RN [4]
RP SEQUENCE OF 1-15 AND 44-50.
RX MEDLINE=76184789; PubMed=1268226;
RA Kistler W.S., Keim P.S., Heinrichson R.L.;
RT "Partial structural analysis of the basic chromosomal protein of rat
RT spermatozoa.";
RN Biochim. Biophys. Acta 427:752-757(1976).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC Sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -!- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
CC DNA-HELIX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46939; CAA87061.1; -
DR PIR: A27128; A27128.
DR PIR: S03997; S03997.
DR INTERPRO: IPR000221; -
DR PFAM: PF00260; protamine.P1.1.
DR PROSITE: PS00048; protamine.P1.1.
DR Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INIT_MER 0 0
FT DISULFID 5 5 INTERCHAIN (WITH CYS-21) (BY SIMILARITY).
FT DISULFID 6 14 BY SIMILARITY.
FT DISULFID 21 21 INTERCHAIN (WITH CYS-5) (BY SIMILARITY).
FT DISULFID 36 36 INTERCHAIN (WITH CYS-36) (BY SIMILARITY).
FT DISULFID 37 47 BY SIMILARITY.
SQ SEQUENCE 50 AA: 6889 MW: B60B9F3D1BB6D978 CRC64;

Query Match 25.6%; Score 63.5; DB 1; Length 50;
Best Local Similarity 28.9%; Pred. No. 1.1;
Matches 11; Conservative 11; Mismatches 11; Indels 5; Gaps 1;

OY 6 QTECCQOCRRHQESDPPOQCYCQRKEICEEEY 43
DB 11 RSRRCRRRRRCRR-----RRRCRRRRRRRCRRRSY 43

RESULT 11
AH12_ARAHY STANDARD: PRT; 626 AA.
ID AH12_ARAHY
AC P43238:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ALLEGREN ARA H 1, CLONE P41B (ARA H 1).
OS Arabis hypogaea (Peasut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Arachis.
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV, FLORUNNER:
RX MEDLINE=96013631; PubMed-7560062.
RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;
RT "Recombinant peanut allergen Ara h I expression and IgE binding in
  RT patients with peanut hypersensitivity.";
  J. Clin. Invest. 96:1715-1721(1995).
  CC -1- SIMILARITY: TO 75 SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
  CC CONVULCIN, CONGLYCININ, ETC.).
  CC
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  CC EMBL: L34402; AAB00861.1; -.
  CC HSSP: P50477; ICAM.
  CC INTERPRO: IPR001113; -.
  CC PFAM: PF00546; Seedstore_7s; 1.
  CC Allergen.
  CC
  SQ SEQUENCE 626 AA; 71345 MW; 1A6BBBA41490D0E3 CRC64;

Query Match 25.4%; Score 63; DB 1; Length 626;
Best Local Similarity 36.1%; Pred. No. 9.5;
Matches 13; Conservative 9; Mismatches 12; Indels 2; Gaps 2;

QY 2 QEDPOTEC-QQCCRCRQESDPPOQYCORCKEI 36
   : : : : : : : : : : : : : : : : : :
Db 31 QKTEPCACRCIQSC-QQEPDCLKKACESCRCRL 65

RESULT 12
CAS2_SHEEP STANDARD; PRT; 223 AA.
ID CAS2_SHEEP STANDARD; PRT; 223 AA.
AC P04654;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ALPHA-S2 CASEIN PRECURSOR.
GN CSNIS2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86104467; PubMed-3002499;
RA Boissard M., Pettissant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
  Biochimie 67:1043-1051(1985).
  CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
  CC CALCIUM PHOSPHATE.
  CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
  CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
  CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  CC EMBL: X03238; CAA26983.1; -.
  CC PIR: A25070; KASHS2.
  CC INTERPRO: IPR001588; -.
  CC PFAM: PF00363; caseins; 2.
  CC PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.

```

```

KM Milk: Phosphorylation; Signal.
FT SIGNAL 1 15
FT CHAIN 16 223
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD_RES 23 23
FT MOD_RES 24 24
FT MOD_RES 25 25
FT MOD_RES 72 72
FT MOD_RES 73 73
FT MOD_RES 74 74
FT VARIANT 64 64
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 24.6%; Score 61; DB 1; Length 223;
Best Local Similarity 26.7%; Pred. No. 6.5;
Matches 12; Conservative 12; Mismatches 19; Indels 2; Gaps 1;

QY 1 NOEDPOTECQCCRCRQESDPPOQYCORCKEI--EEERY 43
   : : : : : : : : : : : : : : : : : :
Db 24 SSEPINISGEIYKQEKNNMHIHRKELCTTSCSEVVRNADDEERY 68

RESULT 13
AP04_MACFA STANDARD; PRT; 429 AA.
ID AP04_MACFA STANDARD; PRT; 429 AA.
AC P33621;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE APOlipoprotein A-IV PRECURSOR (APO-AIV).
GN APOA4.
OS Macaca fascicularis (Crap eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciae;
OC Cercopitheciae; Macaca.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93192330; PubMed-8448212;
RA Osada J., Pocovi M., Nicolosi R.J., Schaefer E.J., Ordovas J.M.;
RT "Nucleotide sequences of the Macaca fascicularis apolipoprotein C-II
  RT and A-IV genes.";
  J. Biol. Chem. 268:1172-1176(1993).
  CC -1- FUNCTION: MAY HAVE A ROLE IN CHYLOMICRONS AND VLDL SECRETION AND
  CC CATABOLISM. REQUIRED FOR EFFICIENT ACTIVATION OF LIPOPROTEIN
  CC LIPASE BY APOC-II; POTENT ACTIVATOR OF LCAT. APOA-IV IS A MAJOR
  CC COMPONENT OF HDL AND CHYLOMICRONS.
  CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
  CC -1- TISSUE SPECIFICITY: SECRETED IN PLASMA.
  CC -1- DOMAIN: NINE OF THE THIRTEEN 22-AMINO ACID TANDEM REPEATS (RACH
  CC MERS) OCCURRING IN THIS SEQUENCE ARE PREDICTED TO BE HIGHLY ALPHA-
  CC HELICAL, AND MANY OF THESE HELICES ARE AMPHIPATHIC. THEY MAY
  CC THEREFORE SERVE AS LIPID-BINDING DOMAINS WITH LECITHIN:CHOLESTEROL
  CC ACYLTRANSFERASE (LCAT) ACTIVATING ABILITIES.
  CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
  CC -----
  CC EMBL: X68361; CAA48421.1; -.
  CC PIR: S29565; S29565.
  CC PIR: S30195; S30195.
  CC HSSP: P02649; INFO.
  CC INTERPRO: IPR000074; -.
  CC PFAM: PF01442; Apolipoprotein; 1.

```

| FW | Plasma | Lipid transport | HDL | Chylomicron | Repeat | Signal |
|----|----------|-----------------|-----------|--|--------|--------|
| FT | SIGNAL | 1 | 20 | BY SIMILARITY. | | |
| FT | CHAIN | 21 | 429 | APOLIPOPROTEIN A-IV. | | |
| FT | DOMAIN | 33 | 330 | 13 X 22 AA APPROXIMATE TANDEM REPEATS. | | |
| FT | REPEAT | 33 | 54 | | | |
| FT | REPEAT | 60 | 81 | | | |
| FT | REPEAT | 82 | 103 | | | |
| FT | REPEAT | 115 | 136 | | | |
| FT | REPEAT | 137 | 158 | | | |
| FT | REPEAT | 159 | 180 | | | |
| FT | REPEAT | 181 | 202 | | | |
| FT | REPEAT | 203 | 224 | | | |
| FT | REPEAT | 225 | 246 | | | |
| FT | REPEAT | 247 | 268 | | | |
| FT | REPEAT | 269 | 286 | | | |
| FT | REPEAT | 287 | 308 | | | |
| FT | REPEAT | 309 | 330 | | | |
| FT | DOMAIN | 372 | 420 | GLU/GLN-RICH. | | |
| SO | SEQUENCE | 429 AA: | 49876 MM: | 30458F51D0DB60C | CRC64: | |

| | | | | | | |
|---------|--|--------------|---------------|------------|-------------|--------|
| | Query Match | 24.6% | Score 61; | DB 1; | Length 422; | |
| | Best Local Similarity | 34.1%; | Pred. No. 11; | | | |
| Matches | 14; | Conservative | 12; | Mismatches | 15; | Indels |
| | | | | | 0; | Gaps |
| | | | | | | 0. |
| Qy | 2 QEDPQTCCGCCRCRCRCEGSDPRDQOYCQRCKEICEEEEE | 42 | | | | |
| | : : : : : | : | | | | |
| Dd | 3776 QREGGQEEQQDEEGEEHQQDDQDEEQDQDEEQDQDEEQDQ | 416 | | | | |
| | | | | | | |
| RESULT | 14 | | | | | |
| ID | ANDR_CANFA | STANDARD: | PRT: | 907 AA. | | |
| AC | O9TTE90: | | | | | |
| DC | 30-MAY-2000 (Rel. 39, Created) | | | | | |
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | | | |
| DE | 30-MAY-2000 (Rel. 39, Last annotation update) | | | | | |
| DR | ANROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR). | | | | | |
| DN | AR OR NR3C4. | | | | | |
| GN | Canis familiaris (dog). | | | | | |
| OS | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | | | |
| OC | Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. | | | | | |
| RN | [1] | | | | | |
| RP | SEQUENCE FROM N.A. | | | | | |
| RA | Lu B., Smock S.L., Castleberry T.A., Owen T.A.: | | | | | |
| RT | "Molecular cloning and functional characterization of the canine | | | | | |
| RL | androgen receptor."; | | | | | |
| RU | Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases. | | | | | |
| CC | -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN | | | | | |
| CC | THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR | | | | | |
| CC | PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES. | | | | | |
| CC | -1- SUBCELLULAR LOCATION: NUCLEAR. | | | | | |
| CC | -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN, | | | | | |
| CC | A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN. | | | | | |
| CC | -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY. | | | | | |
| CC | NR3 SUBFAMILY. | | | | | |
| CC | ----- | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/) | | | | | |
| CC | or send an email to license@isb-sib.ch . | | | | | |
| CC | ----- | | | | | |
| DR | EMBL: AF197950; AAF18084.1; - | | | | | |
| DR | PROSITE: PS00931; NUCLEAR_RECEPTOR; 1. | | | | | |
| KW | Receptor; Transcription regulation; DNA-binding; Nuclear protein; | | | | | |
| KW | Zinc-finger; Steroid-binding. | | | | | |
| FT | DOMAIN 1 545 MODULATING (BY SIMILARITY). | | | | | |
| FT | ZN_BIND 547 612 C4-TYPE ZINC FINGERS (TWO). | | | | | |
| FT | ZN_FING 547 567 C4-TYPE. | | | | | |
| FT | ZN_FING 563 607 C4-TYPE. | | | | | |

| FT | DOMAIN | 678 | 907 | LIGAND-BINDING. |
|----|----------|---------|-----------|-------------------------|
| FT | DOMAIN | 55 | 64 | POLY-GLN. |
| FT | DOMAIN | 70 | 76 | POLY-GLN. |
| FT | DOMAIN | 131 | 134 | POLY-GLN. |
| FT | DOMAIN | 180 | 202 | POLY-GLN. |
| FT | DOMAIN | 329 | 332 | POLY-SER. |
| FT | DOMAIN | 375 | 384 | POLY-PRO. |
| FT | DOMAIN | 399 | 405 | POLY-ALA. |
| 50 | SEQUENCE | 907 AA; | 98726 MW; | C681978780D2338AP CRC64 |

```

Query Match      24.6%  Score 61:  DB 1:  Length 907:
Best Local Similarity  48.1%  Pred. No. 20:
Matches 13:  Conservative  5:  Mismatches  9:  Indels  0:  Gaps  0:

OY      5  PQTCCGCGCGRRCROESDPROOYCOR  31
      1  ||  |:::  ||||  ||||  |  76
Db      50  PGRATLQGGGQGGGQGGGQGGGQGG  76

```

| RESULT | 15 |
|------------|--|
| INVO_CANFA | INVO_CANFA |
| ID | INVO_CANFA |
| AC | PI8174 |
| DT | 01-NOV-1990 (Rel. 16, Created) |
| DT | 01-NOV-1990 (Rel. 16, Last sequence update) |
| DT | 01-FEB-1996 (Rel. 33, Last annotation update) |
| GN | INVOLUCRIN. |
| GN | IVL. |
| OS | Canis familiaris (Dog). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis. |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=90348475; PubMed=2385171; |
| RA | Tseng H., Green H.; |
| RT | "The involucrin genes of pig and dog: comparison of their segments of |
| RL | repeats with those of prosimians and higher primates."; |
| CC | Mol. Biol. Evol. 7:293-302(1990). |
| CC | -I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS |
| CC | IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO |
| CC | MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE |
| CC | FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE. |
| CC | -I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND |
| CC | OTHER STRATIFIED SQUAMOUS EPITHELIA. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaborator |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (see http://www.isb-sdb.ch/announce |
| CC | or send an email to license@sdb-sdb.ch). |
| CC | ----- |
| DR | EMBL; M34442; AAA30853.1; -. |
| DR | INTERPRO; IPR002360; -. |
| DR | PROSITE; PS00795; INVOLUCRIN; 1. |
| KW | keratinocyte; Repeat. |
| QJ | SEQUENCE 285 AA; 33384 MW; DCE1BD8899248BFA CRC64; |

| | | | | |
|-----------------------|--------------|--------------|---------------|------------|
| Query Match | 24.2% | Score 60 | DB 1 | Length 285 |
| Best Local Similarity | 31.7% | Pred. No. 10 | | |
| Matches 13 | Conservative | 9 | Mismatches 19 | Indels 0 |
| Gaps | | | | 0 |

| | | | |
|----|-----|---|-----|
| Oy | 2 | QEDPTECCQCCRCRCRQGSDEPRQOQYCCRRCKEICEEEEE | 42 |
| | 1 | ----- | 11 |
| Db | 142 | QOQEQESQEKLELLEDEQKEELQOQEQEQEQEKCEKHNQE | 182 |
| | 1 | ----- | 11 |

```
Search completed: March 1, 2001, 16:16:47
Job time: 427 sec
```


